STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 151270

TO: Janet Epps-Ford **Location: rem/2c05/2c18** **Art Unit: 1635** Apal 29, 2005 **Case Serial Number: 10/086062**	**From: P. Sheppard** **Location: Remsen Building** **Phone: (571) 272-2529** **sheppard@uspto.gov**

Search Notes

04/28/05

STIC-Biotech/ChemLib 151270 mg

From: Epps-Ford, Janet
Sent: Wednesday, April 20, 2005 5:14 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search.

Application 10/086,062

Please search SEQ ID NO: 4 in all pending and published nucleic acid databases.

Thanks,
Janet L. Epps-Ford, Ph.D.
Art Unit 1635
Mailbox: Remsen 2C18
Office: Remsen 2C05
Phone: 571-272-0757
Fax: 571-273-0757

****
STAFF USE ONLY

Searcher:_____
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed:_____
Searcher Prep/Rev. Time:___
Online Time:_____

****
Type of Search

NA#:_____ AA#:_____
Interference:____ SPDI:___
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure#:_____ Text:___
Inventor:____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other(Specify):_____

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 26, 2005, 06:25:00 ; Search time 1680 Seconds
(without alignments)
865.272 Million cell updates/sec

Title: US-10-086-062-4
Perfect score: 30
Sequence: 1 ctggacccctctcgactcgagagttccgct 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	100.0	30	6	AX358109	AX358109 Sequence
c 2	20.4	68.0	595	9	HUMYQ76F01	AF085912 Homo sapi
3	20.4	68.0	98348	9	AL136103	AL136103 Human DNA
4	20.4	68.0	178965	9	AC012048	AC012048 Homo sapi
c 5	20.2	67.3	23211	3	AC099767	AC099767 Caenorhab
c 6	20	66.7	1503	1	AB038027	AB038027 Vibrio sp
c 7	20	66.7	156165	2	AC021453	AC021453 Homo sapi
c 8	20	66.7	157980	9	AC100852	AC100852 Homo sapi
9	19.6	65.3	257595	2	AC123011	AC123011 Rattus no
c 10	19.4	64.7	11406	4	AF107201	AF107201 Equus cab
11	19.4	64.7	70387	3	AC087075	AC087075 Caenorhab
c 12	19.4	64.7	110000	8	AE016816_6	Continuation (7 of
c 13	19.4	64.7	159969	9	AC104791	AC104791 Homo sapi
c 14	19.4	64.7	189269	2	AP006440	AP006440 Homo sapi
c 15	19.4	64.7	196472	2	AC011862	AC011862 Homo sapi
c 16	19.2	64.0	1500	1	AF118021	AF118021 Vibrio ca
c 17	19.2	64.0	1500	6	AR452360	AR452360 Sequence
c 18	19.2	64.0	1500	6	AX201179	AX201179 Sequence
c 19	19	63.3	93695	8	ATF2009	AL021749 Arabidops
c 20	19	63.3	156806	2	AC021353	AC021353 Homo sapi
21	19	63.3	177652	10	AC140251	AC140251 Mus muscu
c 22	19	63.3	199749	8	ATCHRIV68	AL161572 Arabidops
23	19	63.3	203120	10	AC124991	AC124991 Mus muscu
c 24	19	63.3	205475	2	AC130033	AC130033 Rattus no
c 25	19	63.3	213906	2	AC118634	AC118634 Mus muscu
26	19	63.3	223675	9	AC025031	AC025031 Homo sapi
c 27	19	63.3	232320	10	AC127314	AC127314 Mus muscu
c 28	19	63.3	244974	2	AC111249	AC111249 Rattus no
c 29	18.8	62.7	2481	6	AX827828	AX827828 Sequence
c 30	18.8	62.7	2481	10	RATRT1BA2	M22366 Rattus norv
31	18.8	62.7	5550	1	AB016787	AB016787 Pseudomon
32	18.8	62.7	23107	6	AX695653	AX695653 Sequence
c 33	18.8	62.7	35758	2	AC127539	AC127539 Homo sapi
c 34	18.8	62.7	67145	2	AC124281	AC124281 Homo sapi
c 35	18.8	62.7	69833	2	AC010777	AC010777 Homo sapi
c 36	18.8	62.7	77691	9	AC120118	AC120118 Homo sapi
c 37	18.8	62.7	95241	9	HS398I9	AL023096 Human DNA
c 38	18.8	62.7	100521	9	AC119397	AC119397 Homo sapi
39	18.8	62.7	101721	9	AC067950	AC067950 Homo sapi
40	18.8	62.7	110000	8	CR382131_15	Continuation (16 o
c 41	18.8	62.7	113196	9	HSJ697K14	AL121829 Human DNA
c 42	18.8	62.7	141605	2	AC013732	AC013732 Homo sapi
c 43	18.8	62.7	146606	9	AC068213	AC068213 Homo sapi
c 44	18.8	62.7	148332	9	AC079586	AC079586 Homo sapi
c 45	18.8	62.7	148611	8	AC091680	AC091680 Oryza sat

ALIGNMENTS

RESULT 1
AX358109
LOCUS AX358109 30 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 4 from Patent WO0194394.
ACCESSION AX358109
VERSION AX358109.1 GI:18674856
KEYWORDS .
SOURCE Agrobacterium tumefaciens (Rhizobium radiobacter)
ORGANISM Agrobacterium tumefaciens
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE 1
AUTHORS Jilka,J.M., Hood,E.E. and Howard,J.A.
TITLE Novel plant promoter sequences and methods of use for same
JOURNAL Patent: WO 0194394-A 4 13-DEC-2001;
Prodigene, Inc. (US)
FEATURES Location/Qualifiers
source 1. .30
/organism="Agrobacterium tumefaciens"
/mol_type="unassigned DNA"
/db_xref="taxon:358"
ORIGIN

Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
||||||||||||||||||||||||||||||
Db 1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30

RESULT 2
HUMYQ76F01/c
LOCUS HUMYQ76F01 595 bp mRNA linear PRI 29-AUG-1998
DEFINITION Homo sapiens full length insert cDNA clone YQ76F01.
ACCESSION AF085912
VERSION AF085912.1 GI:3483252
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,
            Marth,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,
            Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,
            Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
            Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behymer,K., Hillier,L.,
            Wilson,R. and Waterston,R.
  TITLE     Full Clone Sequencing of the Longest Available Member from Each
            Unigene Cluster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 595)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:est@watson.wustl.edu

            NOTICE:  This sequence represents the full insert of this cDNA.  No
            attempt has been made to verify whether this corresponds to the
            full-length of the original mRNA from which it was derived.  We
            have tried to obtain double-stranded, or double chemistry sequence
            across the entire clone, but potentially, there are areas in the
            sequence where this level of coverage was not achieved.
            Nevertheless, we are confident of the accuracy of this sequence as
            all regions of low quality, as defined by PHRAP (P. Green, in
            preparation), were visually inspected and edited accordingly.  The
            consensus quality values for this sequence have been submitted
            separately.

            The location of this clone is unknown.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:201721"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS"
     repeat_region   9. .338
                     /rpt_family="L2"
ORIGIN

  Query Match             68.0%;  Score 20.4;  DB 9;  Length 595;
  Best Local Similarity   80.0%;  Pred. No. 2.9e+02;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||| .|||||||| || |||| |||| ||
Db        566 CTGAGCCCCTCTCGGCTGGAGACTTCCTCT 537


RESULT 3
AL136103
LOCUS       AL136103               98348 bp    DNA     linear   PRI 04-JAN-2001
DEFINITION  Human DNA sequence from clone RP1-250B11 on chromosome 10 Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL136103
VERSION     AL136103.24  GI:9581541
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98348)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 28, 2000 this sequence version replaced gi:9501151.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is
            the entire insert of clone RP1-250B11 This sequence has been
            finished according to sequence map criteria as follows.  An attempt
            is made to resolve all sequencing problems, such as compressions
            and repeats, but not necessarily within known annotated repeat
            sequence elements.  Where the sequence is ambiguous, there is an
            annotation using the 'unsure' feature key. This sequence was
            generated from part of bacterial clone contigs of human chromosome
            10, constructed by the Sanger Centre Chromosome 10 Mapping Group.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP1-250B11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .98348
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP1-250B11"
                     /clone_lib="RPCI-1"
     misc_feature    1. .34
                     /note="Single clone region. Reads derived from clone PCR.
                     Assembly consistent with restriction digest."
     repeat_region   2458. .2765
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   3087. .3267
                     /note="MER39b repeat: matches 355. .546 of consensus"
     repeat_region   3269. .3606
                     /note="MER39 repeat: matches 13. .380 of consensus"
     repeat_region   6106. .6337
                     /note="L2 repeat: matches 2435. .2705 of consensus"
     repeat_region   6364. .6662
                     /note="AluSg repeat: matches 1. .298 of consensus"
     misc_feature    complement(8333. .8799)
                     /note="match: GSS: Em:AQ403443"
     misc_feature    complement(8375. .8758)
                     /note="match: GSS: Em:AQ264393"
     misc_feature    8802. .9234
                     /note="match: GSS: Em:AQ150662"
     misc_feature    9268. .9356
                     /note="match: STS: Em:G28019"
     repeat_region   11809. .12092
                     /note="AluSx repeat: matches 1. .284 of consensus"
     repeat_region   12205. .12722
                     /note="L1ME repeat: matches 5285. .5811 of consensus"
     repeat_region   14113. .14171
                     /note="L2 repeat: matches 2692. .2750 of consensus"
     repeat_region   14548. .14846
                     /note="AluY repeat: matches 1. .299 of consensus"
     repeat_region   15883. .16169
                     /note="AluSx repeat: matches 1. .287 of consensus"
     repeat_region   17146. .18557
                     /note="MER52A repeat: matches 1. .1755 of consensus"
     repeat_region   18559. .18713
                     /note="MIR repeat: matches 6. .160 of consensus"
     repeat_region   18955. .19381
                     /note="MER63 repeat: matches 590. .1061 of consensus"
```

```
repeat_region   19384. .19519
                /note="FLAM_C repeat: matches 3. .137 of consensus"
repeat_region   19713. .19828
                /note="MER63 repeat: matches 463. .577 of consensus"
repeat_region   19826. .20117
                /note="MER63 repeat: matches 2. .307 of consensus"
repeat_region   20325. .20391
                /note="MIR repeat: matches 88. .156 of consensus"
repeat_region   20440. .20701
                /note="MIR repeat: matches 2. .248 of consensus"
misc_feature    22321. .22631
                /note="match: STS: Em:HSJ10C4"
misc_feature    25725. .26188
                /note="match: GSS: Em:AQ211002"
misc_feature    25738. .25978
                /note="match: GSS: Em:AQ823705"
misc_feature    26548. .27067
                /note="match: GSS: Em:AQ798605"
misc_feature    26564. .27250
                /note="match: GSS: Em:AQ055423"
misc_feature    27549. .27913
                /note="match: GSS: Em:B87793"
repeat_region   28724. .28952
                /note="L2 repeat: matches 2167. .2416 of consensus"
misc_feature    complement(29565. .30037)
                /note="match: GSS: Em:AQ155973"
misc_feature    30073. .30670
                /note="match: GSS: Em:AQ195052"
misc_feature    complement(31752. .32121)
                /note="match: GSS: Em:AQ099305"
misc_feature    complement(32344. .32928)
                /note="match: GSS: Em:AQ485233"
misc_feature    32930. .33368
                /note="match: GSS: Em:AQ424293"
repeat_region   34554. .34583
                /note="15 copies 2 mer ca 100% conserved"
repeat_region   35059. .35135
                /note="L2 repeat: matches 2417. .2500 of consensus"
misc_feature    complement(35681. .36191)
                /note="match: GSS: Em:B55923"
misc_feature    complement(35777. .36102)
                /note="match: GSS: Em:AQ223744"
misc_feature    complement(35861. .36186)
                /note="match: GSS: Em:AQ803876"
repeat_region   36692. .36800
                /note="MIR repeat: matches 13. .123 of consensus"
repeat_region   40461. .40765
                /note="AluSx repeat: matches 1. .293 of consensus"
repeat_region   41312. .41502
                /note="MIR repeat: matches 8. .192 of consensus"
repeat_region   41785. .41840
                /note="28 copies 2 mer tg 85% conserved"
repeat_region   43022. .43334
                /note="AluSx repeat: matches 3. .312 of consensus"
misc_feature    complement(44477. .45066)
                /note="match: GSS: Em:AQ540223"
misc_feature    complement(44671. .45063)
                /note="match: GSS: Em:AQ337658"
repeat_region   45077. .45157
                /note="MER57-internal repeat: matches 7151. .7230 of
                consensus"
repeat_region   45158. .45304
                /note="L1P repeat: matches 5050. .5197 of consensus"
repeat_region   45284. .45621
                /note="L1P4 repeat: matches 5476. .5812 of consensus"
misc_feature    complement(45941. .46452)
                /note="match: GSS: Em:AQ832143"
repeat_region   48564. .48709
                /note="L2 repeat: matches 2596. .2749 of consensus"
repeat_region   48720. .49082
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   49176. .49629
                /note="MLT1D repeat: matches 1. .502 of consensus"
repeat_region   49784. .49912
                /note="FLAM_A repeat: matches 8. .132 of consensus"
repeat_region   49914. .50164
                /note="MIR repeat: matches 2. .261 of consensus"
repeat_region   52850. .52918
                /note="L2 repeat: matches 2635. .2706 of consensus"
misc_feature    53092. .53613
                /note="match: GSS: Em:AQ800606"
misc_feature    53110. .53608
                /note="match: GSS: Em:AQ172250"
misc_feature    53151. .53604
                /note="match: GSS: Em:AQ170181"
misc_feature    53165. .53586
                /note="match: GSS: Em:AQ545768"
misc_feature    53179. .53355
                /note="match: GSS: Em:AQ898990"
misc_feature    53378. .53834
                /note="match: GSS: Em:AQ898990"
repeat_region   53776. .53958
                /note="MIR repeat: matches 41. .250 of consensus"
misc_feature    complement(53900. .54343)
                /note="match: GSS: Em:AQ400632"
repeat_region   53959. .54203
                /note="AluSq repeat: matches 65. .308 of consensus"
repeat_region   54204. .54216
                /note="MIR repeat: matches 250. .260 of consensus"
misc_feature    54361. .54767
                /note="match: GSS: Em:AQ729256"
repeat_region   54372. .54565
                /note="MIR repeat: matches 13. .261 of consensus"
misc_feature    54440. .54730
                /note="match: GSS: Em:AQ084956"
misc_feature    complement(55008. .55535)
                /note="match: GSS: Em:AQ556262"
misc_feature    55610. .55935
                /note="match: GSS: Em:AQ594863"
repeat_region   56871. .57293
                /note="L1MD repeat: matches 883. .1319 of consensus"
repeat_region   57297. .57602
                /note="AluJo repeat: matches 1. .305 of consensus"

  Query Match             68.0%;  Score 20.4;  DB 9;  Length 98348;
  Best Local Similarity   80.0%;  Pred. No. 2e+02;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              |||  |||||||||  || ||||  ||||  ||
Db      62965 CTGAGCCCCTCTCGGCTGGAGACTTCCTCT 62994


RESULT 4
AC012048
LOCUS       AC012048               178965 bp    DNA     linear   PRI 20-APR-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-43N22, complete sequence.
ACCESSION   AC012048
VERSION     AC012048.11  GI:19744964
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178965)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178965)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1999) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 178965)
  AUTHORS   Smith,D.R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 178965)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   5  (bases 1 to 178965)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   6  (bases 1 to 178965)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Mar 27, 2002 this sequence version replaced gi:15193325.
FEATURES             Location/Qualifiers
     source          1. .178965
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-43N22"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             68.0%;  Score 20.4;  DB 9;  Length 178965;
  Best Local Similarity   80.0%;  Pred. No. 1.9e+02;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              |||  |||||||||  || |||| |||| ||
Db      32913 CTGAGCCCCTCTCGGCTGGAGACTTCCTCT 32942


RESULT 5
AC099767/c
LOCUS       AC099767               23211 bp    DNA     linear   INV 20-NOV-2001
DEFINITION  Caenorhabditis briggsae cosmid G24F01, complete sequence.
ACCESSION   AC099767
VERSION     AC099767.1  GI:17017639
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 23211)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 23211)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 23211)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                   Genome Sequencing Center
                   Department of Genetics, Washington University
                   St. Louis , MO  63110, USA
                   email: rw@nematode.wustl.edu

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .23211
                     /organism="Caenorhabditis briggsae"
                     /mol_type="genomic DNA"
                     /strain="GujArat G16"
                     /db_xref="taxon:6238"
ORIGIN

  Query Match             67.3%;  Score 20.2;  DB 3;  Length 23211;
  Best Local Similarity   88.0%;  Pred. No. 2.7e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 GACCCCTCTCGACTCGAGAGTTCCG 28
              ||  |||||| ||||||||||||||
Db      14462 GATTCCTCTCAACTCGAGAGTTCCG 14438


RESULT 6
AB038027/c
LOCUS       AB038027               1503 bp    DNA     linear   BCT 18-FEB-2000
DEFINITION  Vibrio sp. SG128 gene for 16S rRNA, partial sequence.
ACCESSION   AB038027
VERSION     AB038027.1  GI:7007473
KEYWORDS    16S ribosomal RNA.
SOURCE      Vibrio sp. SG128
  ORGANISM  Vibrio sp. SG128
            Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
            Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 1503)
  AUTHORS   Urakawa,H.
  TITLE     16S rRNA gene of marine bacterium
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 1503)
  AUTHORS   Urakawa,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2000) Hidetoshi Urakawa, Northwestern University,
            Department of Civil Engineering; Technological Institute 2145
            Sheridan Road, Evanston, Illinois 60208-3109, USA
            (E-mail:h-urakawa@nwu.edu, Tel:+1-847-467-5710,
            Fax:+1-847-491-4011)
FEATURES             Location/Qualifiers
     source          1. .1503
                     /organism="Vibrio sp. SG128"
                     /mol_type="genomic DNA"
                     /strain="SG128"
                     /db_xref="taxon:115126"
     rRNA            <1. .>1503
                     /product="16S ribosomal RNA"
ORIGIN

  Query Match             66.7%;  Score 20;  DB 1;  Length 1503;
  Best Local Similarity   82.1%;  Pred. No. 4.1e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCG 28
              |||| ||| || |||| |||||| ||||
Db        231 CTGGGCCCATCCCGACGCGAGAGGTCCG 204


RESULT 7
AC021453/c
LOCUS       AC021453               156165 bp    DNA     linear   HTG 01-APR-2000
DEFINITION  Homo sapiens clone RP11-125C16, WORKING DRAFT SEQUENCE, 14
            unordered pieces.
ACCESSION   AC021453
VERSION     AC021453.3  GI:7382318
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156165)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-125C16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156165)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 1, 2000 this sequence version replaced gi:6721267.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5169
                Center clone name: 125_C_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 145142 bases at least Q40
                Consensus quality: 151009 bases at least Q30
                Consensus quality: 153345 bases at least Q20
                Insert size: 160000; agarose-fp
                Insert size: 154865; sum-of-contigs
                Quality coverage: 4.1 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1836: contig of 1836 bp in length
            *     1837     1936: gap of 100 bp
            *     1937     5796: contig of 3860 bp in length
            *     5797     5896: gap of 100 bp
            *     5897     9797: contig of 3901 bp in length
            *     9798     9897: gap of 100 bp
            *     9898    13919: contig of 4022 bp in length
            *    13920    14019: gap of 100 bp
            *    14020    17400: contig of 3381 bp in length
            *    17401    17500: gap of 100 bp
            *    17501    21253: contig of 3753 bp in length
            *    21254    21353: gap of 100 bp
            *    21354    31096: contig of 9743 bp in length
            *    31097    31196: gap of 100 bp
            *    31197    41465: contig of 10269 bp in length
            *    41466    41565: gap of 100 bp
            *    41566    50901: contig of 9336 bp in length
            *    50902    51001: gap of 100 bp
            *    51002    62041: contig of 11040 bp in length
            *    62042    62141: gap of 100 bp
            *    62142    75247: contig of 13106 bp in length
            *    75248    75347: gap of 100 bp
            *    75348    89874: contig of 14527 bp in length
            *    89875    89974: gap of 100 bp
            *    89975   116362: contig of 26388 bp in length
            *   116363   116462: gap of 100 bp
            *   116463   156165: contig of 39703 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156165
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-125C16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1836
                     /note="assembly_fragment"
     misc_feature    1937. .5796
                     /note="assembly_fragment"
     misc_feature    5897. .9797
                     /note="assembly_fragment"
     misc_feature    9898. .13919
                     /note="assembly_fragment"
     misc_feature    14020. .17400
                     /note="assembly_fragment"
     misc_feature    17501. .21253
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    21354. .31096
                     /note="assembly_fragment"
     misc_feature    31197. .41465
                     /note="assembly_fragment"
     misc_feature    41566. .50901
                     /note="assembly_fragment"
     misc_feature    51002. .62041
                     /note="assembly_fragment"
     misc_feature    62142. .75247
                     /note="assembly_fragment"
     misc_feature    75348. .89874
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    89975. .116362
                     /note="assembly_fragment"
     misc_feature    116463. .156165
                     /note="assembly_fragment"
ORIGIN

  Query Match             66.7%;  Score 20;  DB 2;  Length 156165;
  Best Local Similarity   82.1%;  Pred. No. 2.9e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 TGGACCCCTCTCGACTCGAGAGTTCCGC 29
              |||||| ||||| | || |||||||  ||
Db      11285 TGGACCTCTCTCTATTCAAGAGTTCTGC 11258


RESULT 8
AC100852/c
LOCUS       AC100852              157980 bp    DNA     linear   PRI 29-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-125C16, complete sequence.
ACCESSION   AC100852
VERSION     AC100852.2  GI:22539166
KEYWORDS    HTG.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157980)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-125C16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157980)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 157980)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 29, 2002 this sequence version replaced gi:17048222.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21587
                Center clone name: 125_C_16
            --------------.
FEATURES             Location/Qualifiers
     source          1. .157980
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-125C16"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   62. .173
                     /rpt_family="MIR"
     repeat_region   complement(636. .832)
                     /rpt_family="MIR"
     repeat_region   833. .888
                     /rpt_family="MER3"
     repeat_region   complement(891. .1207)
                     /rpt_family="AluJb"
     repeat_region   1208. .1283
                     /rpt_family="MER3"
     unsure          1232. .1236
                     /note="< 30 qual SNGL region"
     repeat_region   complement(1303. .2370)
                     /rpt_family="L1PA4"
     unsure          1330. .1354
                     /note="<30 qual SNGL region"
     repeat_region   2371. .3213
                     /rpt_family="L1PA4"
     repeat_region   complement(3266. .3408)
                     /rpt_family="Charlie7"
     repeat_region   3506. .3538
                     /rpt_family="L2"
     repeat_region   3539. .3653
                     /rpt_family="MER47A"
     repeat_region   complement(3654. .3926)
                     /rpt_family="AluSc"
     repeat_region   3927. .4141
                     /rpt_family="MER47A"
     repeat_region   4142. .4438
                     /rpt_family="AluY"
     repeat_region   4439. .4473
                     /rpt_family="MER47A"
     repeat_region   4474. .4561
                     /rpt_family="L2"
     repeat_region   complement(4601. .4700)
                     /rpt_family="L2"
     repeat_region   complement(4708. .4845)
                     /rpt_family="MER5A"
     repeat_region   complement(4944. .5104)
                     /rpt_family="MIR"
     repeat_region   complement(5112. .5532)
                     /rpt_family="L2"
     repeat_region   complement(5546. .5822)
                     /rpt_family="AluJb"
     repeat_region   complement(5823. .5999)
                     /rpt_family="L2"
     repeat_region   6000. .6216
                     /rpt_family="AluJb"
     repeat_region   complement(6217. .6292)
                     /rpt_family="L2"
     repeat_region   6447. .6752
                     /rpt_family="AluJo"
     repeat_region   7412. .7472
                     /rpt_family="A-rich"
     repeat_region   7789. .7907
                     /rpt_family="(CGG)n"
     repeat_region   8642. .8683
                     /rpt_family="T-rich"
     repeat_region   8954. .8978
                     /rpt_family="(CCA)n"
     repeat_region   9084. .9123
                     /rpt_family="(CAG)n"
     repeat_region   10990. .11017
                     /rpt_family="AT_rich"
     repeat_region   11225. .11326
                     /rpt_family="L2"
     repeat_region   complement(11372. .11504)
                     /rpt_family="MLT1B"
```

```
     repeat_region   13034. .13209
                     /rpt_family="MIR"
     repeat_region   complement(15480. .15590)
                     /rpt_family="L2"
     repeat_region   15703. .15988
                     /rpt_family="AluSq"
     repeat_region   complement(16005. .16318)
                     /rpt_family="L2"
     repeat_region   16412. .16555
                     /rpt_family="MER106B"
     repeat_region   16825. .17099
                     /rpt_family="AluJb"
     repeat_region   17241. .17261
                     /rpt_family="AT_rich"
     repeat_region   17280. .17310
                     /rpt_family="(TTTG)n"
     repeat_region   complement(17366. .17658)
                     /rpt_family="AluSc"
     repeat_region   17707. .17808
                     /rpt_family="MIR"
     repeat_region   18250. .18277
                     /rpt_family="AT_rich"
     repeat_region   complement(18576. .18864)
                     /rpt_family="MER33"
     repeat_region   complement(18926. .19077)
                     /rpt_family="AluJo"
     repeat_region   20786. .20952
                     /rpt_family="AluJo"
     repeat_region   20962. .21087
                     /rpt_family="MIR"
     repeat_region   complement(21088. .21390)
                     /rpt_family="AluSc"
     repeat_region   complement(21419. .21646)
                     /rpt_family="AluJo"
     repeat_region   21647. .21715
                     /rpt_family="MIR"
     repeat_region   complement(21969. .22223)
                     /rpt_family="L1ME4A"
     repeat_region   complement(22288. .22411)
                     /rpt_family="MIR3"
     repeat_region   complement(22950. .23256)
                     /rpt_family="AluSp"
     repeat_region   24607. .24789
                     /rpt_family="FAM"
     repeat_region   24846. .25144
                     /rpt_family="AluJb"
     repeat_region   25229. .25254
                     /rpt_family="AT_rich"
     repeat_region   25700. .26104

  Query Match             66.7%;  Score 20;  DB 9;  Length 157980;
  Best Local Similarity   82.1%;  Pred. No. 2.9e+02;
  Matches   23;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 TGGACCCCTCTCGACTCGAGAGTTCCGC 29
              |||||| ||||| | || ||||||| ||
Db     130676 TGGACCTCTCTCTATTCAAGAGTTCTGC 130649


RESULT 9
AC123011
LOCUS       AC123011              257595 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-92I24, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC123011
VERSION     AC123011.3  GI:23665280
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 257595)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 257595)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 257595)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 10, 2002 this sequence version replaced gi:21909149.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
```

```
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYWD
                Center clone name: CH230-92I24
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 236342 bases at least Q40
                Consensus quality: 239870 bases at least Q30
                Consensus quality: 242005 bases at least Q20
                Estimated insert size: 243903; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   250677: contig of 250677 bp in length
            *   250678   250777: gap of unknown length
            *   250778   252307: contig of 1530 bp in length
            *   252308   252407: gap of unknown length
            *   252408   253577: contig of 1170 bp in length
            *   253578   253677: gap of unknown length
            *   253678   257595: contig of 3918 bp in length.
FEATURES             Location/Qualifiers
     source          1. .257595
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-92I24"
     misc_feature    complement(246343. .247143)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH295371"
     misc_feature    complement(246398. .247079)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH295371"
ORIGIN

  Query Match             65.3%;  Score 19.6;  DB 2;  Length 257595;
  Best Local Similarity   84.6%;  Pred. No. 4.2e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTC 26
              ||||||| |||| ||| | |||||||
Db     151391 CTGGACCTCTCTTGACCCTAGAGTTC 151416


RESULT 10
AF107201/c
LOCUS       AF107201               11406 bp    DNA     linear   MAM 13-DEC-1998
DEFINITION  Equus caballus beta-lactoglobulin II (BLG) gene, complete cds.
ACCESSION   AF107201
VERSION     AF107201.1  GI:4008111
KEYWORDS    .
SOURCE      Equus caballus (horse)
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 11406)
  AUTHORS   Masel,A.M., Brandon,R.B. and Bell,T.K.
  TITLE     Nucleotide Sequence of the Equine Beta-Lactoglobulin Gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11406)
  AUTHORS   Masel,A.M., Brandon,R.B. and Bell,T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1998) Australian Equine Blood Typing Research
            Laboratory, University of Queensland, St. Lucia, Brisbane,
            Queensland 4072, Australia
FEATURES             Location/Qualifiers
     source          1. .11406
                     /organism="Equus caballus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9796"
     gene            <4472. .>8679
                     /gene="BLG"
     mRNA            join(<4472. .4567,4956. .5095,5767. .5840,7367. .7480,
                     8302. .8406,8663. .>8679)
                     /gene="BLG"
                     /product="beta-lactoglobulin II"
     CDS             join(4472. .4567,4956. .5095,5767. .5840,7367. .7480,
                     8302. .8406,8663. .8679)
                     /gene="BLG"
                     /codon_start=1
                     /product="beta-lactoglobulin II"
                     /protein_id="AAC95385.1"
                     /db_xref="GI:4008112"
                     /translation="MKCLLLALGLSLMCGNQATDIPQTMQDLDLQEVAGRWHSVAMVA
                     SDISLLDSESVPLRVYVEELRPTPEGNLEIILREGANHACVERNIVAQKTEDPAVFTV
                     NYQGERKISVLDTDYAHYMFFCVGPPLPSAEHGMVCQYLARTQKVDEEVMEKFSRALQ
                     PLPGRVQIVQDPSGGQERCGF"
ORIGIN

  Query Match             64.7%;  Score 19.4;  DB 4;  Length 11406;
  Best Local Similarity   79.3%;  Pred. No. 6.5e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGC 29
              | || ||||||| || |||||||||| | |
Db       6919 CGGGGCCCCTCTTGAGTCGAGAGTTACAC 6891


RESULT 11
AC087075
LOCUS       AC087075               70387 bp    DNA     linear   INV 05-DEC-2000
DEFINITION  Caenorhabditis briggsae cosmid CB023K10, complete sequence.
ACCESSION   AC087075
VERSION     AC087075.1  GI:11545986
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 70387)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70387)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                     Genome Sequencing Center
                     Department of Genetics, Washington University,
                     St. Louis, MO 63110, USA
                     e-mail: jspieth@watson.wustl.edu


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
```

```
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-181K12"
                     /clone_lib="RPCI-11"
     repeat_region   191. .463
                     /rpt_family="MIR"
     repeat_region   1855. .2217
                     /rpt_family="MaLR"
     repeat_region   2333. .2362
                     /rpt_family="AT_rich"
     repeat_region   2471. .2690
                     /rpt_family="MIR"
     repeat_region   3143. .3169
                     /rpt_family="(A)n"
     repeat_region   3296. .3316
                     /rpt_family="AT_rich"
     repeat_region   3304. .3462
                     /rpt_family="Alu"
     repeat_region   3589. .3609
                     /rpt_family="AT_rich"
     repeat_region   3984. .4033
                     /rpt_family="ERVL"
     repeat_region   4019. .4111
                     /rpt_family="(TA)n"
     repeat_region   4087. .4495
                     /rpt_family="ERVL"
     repeat_region   5195. .5483
                     /rpt_family="Alu"
     repeat_region   5374. .5402
                     /rpt_family="(GGATG)n"
     repeat_region   5897. .5927
                     /rpt_family="(TTTTG)n"
     repeat_region   5906. .6217
                     /rpt_family="Alu"
     misc_feature    6577. .6760
                     /note="similar to Homo sapiens EST AW843469
                     (NID:g7937452)"
     misc_feature    7129. .7131
                     /note="similar to Homo sapiens EST AW843469
                     (NID:g7937452)"
     repeat_region   7314. .7393
                     /rpt_family="MIR"
     repeat_region   7394. .7415
                     /rpt_family="(T)n"
     repeat_region   7700. .7766
                     /rpt_family="MIR"
     repeat_region   9154. .9438
                     /rpt_family="Alu"
     repeat_region   9437. .9466
                     /rpt_family="(TAA)n"
     repeat_region   10074. .10204
                     /rpt_family="MIR"
     repeat_region   10351. .10812
                     /rpt_family="MaLR"
     repeat_region   11128. .11343
                     /rpt_family="Alu"
     repeat_region   11414. .11626
                     /rpt_family="MER1_type"
     repeat_region   13354. .13449
                     /rpt_family="MER1_type"
     repeat_region   13450. .13721
                     /rpt_family="Alu"
     repeat_region   13722. .13784
                     /rpt_family="MER1_type"
     repeat_region   13813. .13852
                     /rpt_family="CT-rich"
     repeat_region   15896. .15941
                     /rpt_family="(TG)n"
     repeat_region   16404. .16916
                     /rpt_family="L2"
     repeat_region   16918. .16939
                     /rpt_family="(T)n"
     repeat_region   18098. .18220
                     /rpt_family="L1"
     repeat_region   18250. .18539
                     /rpt_family="Alu"
     repeat_region   18545. .18594
                     /rpt_family="(A)n"
     misc_feature    18829. .19205
                     /note="match to EST AA743144 (NID:g2782650) ny20h09.s1"
     repeat_region   19565. .19869
                     /rpt_family="Alu"
     repeat_region   20046. .20785
                     /rpt_family="ERV1"
     repeat_region   21452. .21689
                     /rpt_family="L1"
     repeat_region   21764. .22061
                     /rpt_family="Alu"
     misc_feature    21780. .22172
                     /note="match to EST AA554485 (NID:g2325024) ni36b03.s1"
     repeat_region   22691. .22990
                     /rpt_family="Alu"
     repeat_region   23268. .23312
                     /rpt_family="(TTTA)n"
     repeat_region   23286. .23596
                     /rpt_family="Alu"
     repeat_region   23637. .24312
                     /rpt_family="MER2_type"
     repeat_region   24463. .24756
                     /rpt_family="L2"
     repeat_region   24798. .25023
                     /rpt_family="MER2_type"
     repeat_region   25050. .25411
                     /rpt_family="ERVL"
     repeat_region   25377. .25406
                     /rpt_family="(CAAA)n"
     repeat_region   25412. .25733
                     /rpt_family="Alu"
     repeat_region   25532. .25561
                     /rpt_family="(A)n"

  Query Match             64.7%;  Score 19.4;  DB 9;  Length 159969;
  Best Local Similarity   79.3%;  Pred. No. 5.3e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 TGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              |||||  |||||| ||| ||| |||| ||
Db     120906 TGGACTACTCTCGTCTCCAGACTTCCTCT 120878


RESULT 14
AP006440/c
LOCUS       AP006440              189269 bp    DNA     linear   HTG 21-MAY-2003
DEFINITION  Homo sapiens chromosome 11 clone RP11-368I20 map 11q, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AP006440
VERSION     AP006440.1  GI:30962586
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA of 11q
  JOURNAL   Published Only in Database (2003)
REFERENCE   2  (bases 1 to 189269)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
```

```
          once, or longer because we provide a small overlap between
          neighboring submissions.
FEATURES             Location/Qualifiers
     source          1. .70387
                     /organism="Caenorhabditis briggsae"
                     /mol_type="genomic DNA"
                     /strain="GujArat G16"
                     /db_xref="taxon:6238"
                     /clone="CB023K10"
     tRNA            23139. .23209
                     /product="tRNA-Gly"
                     /note="codon recognized: GGC"
     tRNA            complement(36569. .36640)
                     /product="tRNA-Glu"
                     /note="codon recognized: GAG"
     tRNA            37437. .37508
                     /product="tRNA-Gln"
                     /note="codon recognized: CAA"
ORIGIN

  Query Match             64.7%;  Score 19.4;  DB 3;  Length 70387;
  Best Local Similarity   79.3%;  Pred. No. 5.7e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 TGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              |||  | ||||||| |||| ||||| |||
Db       6252 TGGCACGCTCTCGATTCGAAAGTTCAGCT 6280


RESULT 12
AE016816_6/c
WPCOMMENT
  Sequence split into 9 fragments  LOCUS AE016816 Accession AE016816
     Fragment Name      Begin     End
     AE016816_0             1    110000
     AE016816_1        100001    210000
     AE016816_2        200001    310000
     AE016816_3        300001    410000
     AE016816_4        400001    510000
     AE016816_5        500001    610000
     AE016816_6        600001    710000
     AE016816_7        700001    810000
     AE016816_8        800001    907057
Continuation (7 of 9) of AE016816 from base 600001 (AE016816 Eremothecium gossypii ATCC

  Query Match             64.7%;  Score 19.4;  DB 8;  Length 110000;
  Best Local Similarity   79.3%;  Pred. No. 5.5e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGC 29
              |||||  || |||| |||||| |||| |
Db      20976 CTGGACTGCTGTCGAATCGAGACTTCCAC 20948


RESULT 13
AC104791/c
LOCUS       AC104791              159969 bp    DNA     linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-181K12 from 4, complete sequence.
ACCESSION   AC104791 AC032008
VERSION     AC104791.3  GI:18482313
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159969)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 159969)
  AUTHORS   Isak,A., Meyer,R. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-181K12
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 159969)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 159969)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 159969)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 3, 2002 this sequence version replaced gi:18042374.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0181K12
                Drafting Center: WIBR
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-292M9; the clone sequenced
            to the right is RP11-203B7.  Actual start of this clone is at base
            position 1 of RP11-181K12; actual end is at base position 159969 of
            RP11-181K12.

            A transposon has been identified in the vector of this clone.

            The sequence of AC032008 has been incorporated into AC104791.
FEATURES             Location/Qualifiers
     source          1. .159969
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
COMMENT     ------------- Genome Center
              Center: RIKEN Genomic Sciences Center(GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@gsc.riken.go.jp
            ------------- Project Information
              Center project name: HumDraft11
              Center clone name: RP11-368I20
            ------------- Summary Statistics
              Sequencing vector: PCR products; 100% of reads
              Chemistry: Dye-terminator ET-amersham; 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 188057 bases at least Q40
              Consensus quality: 188482 bases at least Q30
              Consensus quality: 188825 bases at least Q20
              Insert size: 188969; sum-of-contigs
              Quality coverage: 12.20x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            4 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     70266 contig of  70266 bp in length
                70367    130873 contig of  60507 bp in length
               130974    188629 contig of  57656 bp in length
               188730    189269 contig of    540 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    70266: contig of 70266 bp in length
            *   70267    70366: gap of 100 bp
            *   70367   130873: contig of 60507 bp in length
            *  130874   130973: gap of 100 bp
            *  130974   188629: contig of 57656 bp in length
            *  188630   188729: gap of 100 bp
            *  188730   189269: contig of 540 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189269
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-368I20"
     misc_feature    1. .70266
                     /note="assembly_fragment"
     misc_feature    70367. .130873
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    130974. .188629
                     /note="assembly_fragment"
     misc_feature    188730. .189269
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
ORIGIN

  Query Match             64.7%;  Score 19.4;  DB 2;  Length 189269;
  Best Local Similarity   79.3%;  Pred. No. 5.3e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 TGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              |||||  ||||||  ||| |||  |||| ||
Db     161287 TGGACTACTCTCGTCTCCAGACTTCCTCT 161259


RESULT 15
AC011862/c
```

```
LOCUS       AC011862              196472 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-15K8, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC011862
VERSION     AC011862.4  GI:7144913
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196472)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-15K8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196472)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6980935.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
              Center project name: L3484
              Center clone name: 15_K_8
            --------------.
            * NOTE: This record contains 217 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      883: contig of 883 bp in length
            *     884      983: gap of 100 bp
            *     984     1877: contig of 894 bp in length
            *    1878     1977: gap of 100 bp
            *    1978     2866: contig of 889 bp in length
            *    2867     2966: gap of 100 bp
            *    2967     3829: contig of 863 bp in length
            *    3830     3929: gap of 100 bp
            *    3930     4831: contig of 902 bp in length
            *    4832     4931: gap of 100 bp
            *    4932     5790: contig of 859 bp in length
            *    5791     5890: gap of 100 bp
            *    5891     6752: contig of 862 bp in length
            *    6753     6852: gap of 100 bp
            *    6853     7749: contig of 897 bp in length
            *    7750     7849: gap of 100 bp
            *    7850     8754: contig of 905 bp in length
            *    8755     8854: gap of 100 bp
```

```
*     8855     9749: contig of 895 bp in length
*     9750     9849: gap of 100 bp
*     9850    10738: contig of 889 bp in length
*    10739    10838: gap of 100 bp
*    10839    11725: contig of 887 bp in length
*    11726    11825: gap of 100 bp
*    11826    12707: contig of 882 bp in length
*    12708    12807: gap of 100 bp
*    12808    13697: contig of 890 bp in length
*    13698    13797: gap of 100 bp
*    13798    14668: contig of 871 bp in length
*    14669    14768: gap of 100 bp
*    14769    15648: contig of 880 bp in length
*    15649    15748: gap of 100 bp
*    15749    16622: contig of 874 bp in length
*    16623    16722: gap of 100 bp
*    16723    17587: contig of 865 bp in length
*    17588    17687: gap of 100 bp
*    17688    18558: contig of 871 bp in length
*    18559    18658: gap of 100 bp
*    18659    19534: contig of 876 bp in length
*    19535    19634: gap of 100 bp
*    19635    20510: contig of 876 bp in length
*    20511    20610: gap of 100 bp
*    20611    21517: contig of 907 bp in length
*    21518    21617: gap of 100 bp
*    21618    22499: contig of 882 bp in length
*    22500    22599: gap of 100 bp
*    22600    23480: contig of 881 bp in length
*    23481    23580: gap of 100 bp
*    23581    24444: contig of 864 bp in length
*    24445    24544: gap of 100 bp
*    24545    25479: contig of 935 bp in length
*    25480    25579: gap of 100 bp
*    25580    26492: contig of 913 bp in length
*    26493    26592: gap of 100 bp
*    26593    27454: contig of 862 bp in length
*    27455    27554: gap of 100 bp
*    27555    28431: contig of 877 bp in length
*    28432    28531: gap of 100 bp
*    28532    29495: contig of 964 bp in length
*    29496    29595: gap of 100 bp
*    29596    30489: contig of 894 bp in length
*    30490    30589: gap of 100 bp
*    30590    31435: contig of 846 bp in length
*    31436    31535: gap of 100 bp
*    31536    32425: contig of 890 bp in length
*    32426    32525: gap of 100 bp
*    32526    33401: contig of 876 bp in length
*    33402    33501: gap of 100 bp
*    33502    34376: contig of 875 bp in length
*    34377    34476: gap of 100 bp
*    34477    35355: contig of 879 bp in length
*    35356    35455: gap of 100 bp
*    35456    36370: contig of 915 bp in length
*    36371    36470: gap of 100 bp
*    36471    37358: contig of 888 bp in length
*    37359    37458: gap of 100 bp
*    37459    38332: contig of 874 bp in length
*    38333    38432: gap of 100 bp
*    38433    39285: contig of 853 bp in length
*    39286    39385: gap of 100 bp
*    39386    40317: contig of 932 bp in length
*    40318    40417: gap of 100 bp
*    40418    41314: contig of 897 bp in length
*    41315    41414: gap of 100 bp
*    41415    42297: contig of 883 bp in length
*    42298    42397: gap of 100 bp
*    42398    43270: contig of 873 bp in length
*    43271    43370: gap of 100 bp
*    43371    44268: contig of 898 bp in length
*    44269    44368: gap of 100 bp
*    44369    45233: contig of 865 bp in length
*    45234    45333: gap of 100 bp
*    45334    46202: contig of 869 bp in length
*    46203    46302: gap of 100 bp
*    46303    47193: contig of 891 bp in length
*    47194    47293: gap of 100 bp
*    47294    48182: contig of 889 bp in length
*    48183    48282: gap of 100 bp
*    48283    49146: contig of 864 bp in length
*    49147    49246: gap of 100 bp
*    49247    50116: contig of 870 bp in length
*    50117    50216: gap of 100 bp
*    50217    51112: contig of 896 bp in length
*    51113    51212: gap of 100 bp
*    51213    52102: contig of 890 bp in length
*    52103    52202: gap of 100 bp
*    52203    53064: contig of 862 bp in length
*    53065    53164: gap of 100 bp
*    53165    54024: contig of 860 bp in length
*    54025    54124: gap of 100 bp
*    54125    54957: contig of 833 bp in length
*    54958    55057: gap of 100 bp
*    55058    55938: contig of 881 bp in length
*    55939    56038: gap of 100 bp
*    56039    56932: contig of 894 bp in length
*    56933    57032: gap of 100 bp
*    57033    57916: contig of 884 bp in length
*    57917    58016: gap of 100 bp
*    58017    58875: contig of 859 bp in length
*    58876    58975: gap of 100 bp
*    58976    59844: contig of 869 bp in length
*    59845    59944: gap of 100 bp
*    59945    60795: contig of 851 bp in length
*    60796    60895: gap of 100 bp
*    60896    61774: contig of 879 bp in length
*    61775    61874: gap of 100 bp
*    61875    62739: contig of 865 bp in length
*    62740    62839: gap of 100 bp
*    62840    63807: contig of 968 bp in length
*    63808    63907: gap of 100 bp
*    63908    64783: contig of 876 bp in length
*    64784    64883: gap of 100 bp
*    64884    65767: contig of 884 bp in length
*    65768    65867: gap of 100 bp
*    65868    66765: contig of 898 bp in length
*    66766    66865: gap of 100 bp
*    66866    67730: contig of 865 bp in length
*    67731    67830: gap of 100 bp
*    67831    68721: contig of 891 bp in length
*    68722    68821: gap of 100 bp
*    68822    69711: contig of 890 bp in length
*    69712    69811: gap of 100 bp
*    69812    70699: contig of 888 bp in length
*    70700    70799: gap of 100 bp
*    70800    71679: contig of 880 bp in length

  Query Match           64.7%;  Score 19.4;  DB 2;  Length 196472;
  Best Local Similarity   79.3%;  Pred. No. 5.3e+02;
  Matches   23;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGC 29
              |||||||||| | || |||||||| |
Db       6946 CTGGACCCCTGGTGTCTGGAGAGTTCCCC 6918


Search completed: April 26, 2005, 11:39:38
Job time : 1686 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: April 26, 2005, 06:19:21 ; Search time 427 Seconds
(without alignments)
415.907 Million cell updates/sec

Title: US-10-086-062-4
Perfect score: 30
Sequence: 1 ctggacccctctcgactcgagagttccgct 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	30	100.0	30	6	AAD24069	Aad24069 Maize eng
c	2	19.2	64.0	1500	5	AAS11024	Aas11024 Vibrio ch
c	3	19	63.3	121124	12	ADQ97107	Adq97107 Mouse can
c	4	18.8	62.7	2481	10	ABT41766	Abt41766 Toxicity
	5	18.8	62.7	23107	9	ADA02762	Ada02762 Human RUN
	6	18.8	62.7	23107	10	ADB72500	Adb72500 Human Run
	7	18.8	62.7	23107	10	ADC85242	Adc85242 Human Run
	8	18.8	62.7	23107	12	ADM74357	Adm74357 Human car
c	9	18.6	62.0	3210	2	AAZ27624	Aaz27624 Plasmid S
c	10	18.6	62.0	3459	2	AAZ27623	Aaz27623 Plasmid S
	11	18.4	61.3	319608	3	AAH51601	Aah51601 Human chr
	12	18.4	61.3	319608	5	AAS09301	Aas09301 Human sch
	13	18.2	60.7	340449	8	AAL52198	Aal52198 Human sec
	14	18	60.0	42325	10	ADB74382	Adb74382 Mycobacte
c	15	18	60.0	110000	11	ADM27081_10	Continuation (11 o
	16	17.8	59.3	165	10	ADD49400	Add49400 Human lun
	17	17.8	59.3	254	12	ADG00322	Adg00322 Nicotiana
	18	17.8	59.3	369	11	ABD11207	Abd11207 Pseudomon
c	19	17.8	59.3	426	10	ADD49385	Add49385 Human lun
c	20	17.8	59.3	449	10	ADD49343	Add49343 Human lun
	21	17.8	59.3	449	10	ADD49294	Add49294 Human lun
c	22	17.8	59.3	986	6	ABQ46720	Abq46720 Oligonucl
	23	17.8	59.3	986	6	ABQ46721	Abq46721 Oligonucl
	24	17.8	59.3	1553	12	ADL12873	Adl12873 Human ste
c	25	17.8	59.3	1603	4	AAK77077	Aak77077 Human imm
	26	17.8	59.3	1731	11	ACN91401	Acn91401 Breast ca
c	27	17.8	59.3	2780	13	ADR25834	Adr25834 Breast ca
c	28	17.8	59.3	152141	8	ACA64961	Aca64961 Human BCR
	29	17.8	59.3	153995	13	ABD33534	Abd33534 Murine ca
	30	17.6	58.7	246	12	ADO52469	Ado52469 Human met
c	31	17.6	58.7	8119	3	AAZ35392	Aaz35392 Maize sta
	32	17.6	58.7	49243	4	ABL03188	Abl03188 Drosophil
c	33	17.4	58.0	65	6	ABN51822	Abn51822 Mouse spl
c	34	17.4	58.0	300	2	AAZ13036	Aaz13036 Human gen
c	35	17.4	58.0	300	2	AAX98464	Aax98464 Human can
c	36	17.4	58.0	411	4	AAI85652	Aai85652 Human pol
c	37	17.4	58.0	460	9	ACH41661	Ach41661 Human foe
	38	17.4	58.0	594	12	ACH74968	Ach74968 Human gen
c	39	17.4	58.0	699	2	AAZ15926	Aaz15926 Human gen
c	40	17.4	58.0	729	12	ADO63441	Ado63441 Transcrip
c	41	17.4	58.0	778	6	ABK30430	Abk30430 Human G-p
	42	17.4	58.0	936	13	ADR92877	Adr92877 Novel S.
	43	17.4	58.0	949	13	ADS50577	Ads50577 Bacterial
c	44	17.4	58.0	1210	4	AAS41038	Aas41038 cDNA enco
c	45	17.4	58.0	1210	4	AAS34819	Aas34819 cDNA enco

ALIGNMENTS

RESULT 1
AAD24069
ID AAD24069 standard; DNA; 30 BP.
XX
AC AAD24069;
XX
DT 09-APR-2002 (first entry)
XX
DE Maize engineered Ubi-1 promoter heat shock element #3.
XX
KW Gene expression; maize; ubiquitin promoter; Ubi-1; HSE;
KW heat shock element; agronomic gene; ds.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .15
FT /*tag= a
FT /note= "5' heat shock element"
FT misc_feature 16. .30
FT /*tag= b
FT /note= "3' heat shock element"
XX
PN WO200194394-A2.
XX
PD 13-DEC-2001.
XX
PF 08-JUN-2001; 2001WO-US018689.
XX
PR 09-JUN-2000; 2000US-00590558.
XX
PA (PROD-) PRODIGENE INC.
XX
PI Jilka JM, Hood EE, Howard JA;
XX
DR WPI; 2002-122117/16.
XX
PT New promoter sequences for causing expression of a structural gene
PT especially agronomic gene or open reading frame in a plant cell,
PT comprises engineered versions of the maize ubiquitin promoter.
XX
PS Claim 6; Page 54; 68pp; English.
XX

```
CC   The invention relates to a promoter sequence capable of directing
CC   expression of a nucleotide sequence in a plant cell, comprising maize
CC   ubiquitin (Ubi-1) promoter sequence with a modification so that it does
CC   not include two overlapping heat shock elements (HSE) or it directs
CC   expression to increase the endosperm/embryo expression ratio of the
CC   protein when compared to the ratio from a wild-type ubiquitin promoter.
CC   The modified Ubi-1 promoter comprises a deletion of 3', 5' or both HSEs,
CC   two non-overlapping/adjacent HSEs, replacement of HSEs with a trimer of a
CC   seed specific element from the promoter of pea lectin gene Ps1 , or
CC   insertion of a transcription factor binding site in the HSE region. An
CC   expression construct comprising modified Ubi-1 promoter is useful for
CC   causing expression of a structural gene (agronomic genes) or open reading
CC   frame in a plant cell. The modified Ubi-1 promoter increases expression
CC   levels beyond those observed with native ubiquitin promoter. The present
CC   sequence is maize engineered Ubi-1 promoter with heat shock elements
CC   adjacently placed. Note: The present sequence is also shown in claim 26,
CC   page-56 of the specification. However, this sequence has an additional
CC   nucleotide at the 3' end
XX
SQ   Sequence 30 BP; 4 A; 12 C; 7 G; 7 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 30;  DB 6;  Length 30;
  Best Local Similarity  100.0%;  Pred. No. 0.0012;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||||||||||||||||||||||||||||||
Db          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30


RESULT 2
AAS11024/c
ID   AAS11024 standard; DNA; 1500 BP.
XX
AC   AAS11024;
XX
DT   11-SEP-2003  (revised)
DT   24-OCT-2001  (first entry)
XX
DE   Vibrio cholera 16s ribosomal RNA gene.
XX
KW   Antisense; bacterial 16s ribosomal RNA; rRNA; bacterial infection; human;
KW   food grain supplement; livestock; poultry; therapeutic; ds.
XX
OS   Vibrio cholerae.
XX
PN   WO200142457-A2.
XX
PD   14-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-US042391.
XX
PR   29-NOV-1999;   99US-0168150P.
XX
PA   (AVIB-) AVI BIOPHARMA INC.
XX
PI   Iversen PL;
XX
DR   WPI; 2001-457295/49.
XX
PT   Antibacterial compound, useful for treating bacterial infections and as
PT   livestock and poultry food supplement, comprises antisense
PT   oligonucleotides complementary to bacterial 16S and 23S rRNA.
XX
PS   Disclosure; Page; 62pp; English.
XX
CC   AAS11021-AAS11034 represent the coding sequences of bacterial 16s
CC   ribosomal RNA (rRNA) genes. The sequences were used to design anti-
CC   bacterial compounds comprising substantially uncharged antisense
CC   oligomers containing 8-40 nucleotide subunits, including a targeting
CC   nucleic acid sequence at least 10 nucleotides in length which is
CC   complementary to a bacterial 16S or 23S rRNA nucleic acid sequence. The
CC   antisense oligomers are used for treating a bacterial infection in a
CC   human or a mammalian animal produced by Escherichia coli, Salmonella
CC   typhimurium, Pseudomonas aeruginosa, Vibrio cholera, Neisseria
CC   gonorrhoea, Helicobacter pylori, Bartonella henselae, Haemophilus
CC   influenza, Shigella dysenterae, Staphylococcus aureus, Mycobacterium
CC   tuberculosis, Streptococcus pneumoniae, Treponema palladium and Chlamydia
CC   trachomatis. The antibacterial compound may be used as a food grain
CC   supplement in livestock and poultry food composition. Note: The present
CC   sequence is not shown in the specification but has been accessed from
CC   GenBank using the appropriate accession number given in the
CC   specification. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ   Sequence 1500 BP; 376 A; 326 C; 482 G; 312 T; 0 U; 4 Other;

  Query Match            64.0%;  Score 19.2;  DB 5;  Length 1500;
  Best Local Similarity  75.0%;  Pred. No. 1.1e+02;
  Matches   21;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCG 28
              |||| ||| || |||| |||:|: ||||
Db        212 CTGGGCCCATCCCGACGCGARARGTCCG 185


RESULT 3
ADQ97107/c
ID   ADQ97107 standard; DNA; 121124 BP.
XX
AC   ADQ97107;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD08-002, SEQ ID 83.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 83; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 121124 BP; 32972 A; 25314 C; 26641 G; 35451 T; 0 U; 746 Other;

  Query Match            63.3%;  Score 19;  DB 12;  Length 121124;
  Best Local Similarity  81.5%;  Pred. No. 1.9e+02;
  Matches   22;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GACCCCTCTCGACTCGAGAGTTCCGCT 30
              | |||||||||| |  ||||||| |||
Db      88764 GCCCCCTCTCGATTACAGAGTTCAGCT 88738
```

```
RESULT 4
ABT41766/c
ID   ABT41766 standard; DNA; 2481 BP.
XX
AC   ABT41766;
XX
DT   26-JUN-2003  (first entry)
XX
DE   Toxicity modelling related rat gene SEQ ID No 1468.
XX
KW   Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW   database; drug screening; toxicity assay; rat; ds.
XX
OS   Rattus norvegicus.
XX
PN   WO200295000-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-US016173.
XX
PR   22-MAY-2001; 2001US-0292335P.
PR   13-JUN-2001; 2001US-0297523P.
PR   19-JUN-2001; 2001US-0298925P.
PR   10-JUL-2001; 2001US-0303807P.
PR   10-JUL-2001; 2001US-0303808P.
PR   10-JUL-2001; 2001US-0303810P.
PR   28-AUG-2001; 2001US-0315047P.
PR   27-SEP-2001; 2001US-0324928P.
PR   22-OCT-2001; 2001US-0330462P.
PR   01-NOV-2001; 2001US-0330867P.
PR   21-NOV-2001; 2001US-0331805P.
PR   06-DEC-2001; 2001US-0336144P.
PR   19-DEC-2001; 2001US-0340873P.
PR   21-FEB-2002; 2002US-0357842P.
PR   21-FEB-2002; 2002US-0357843P.
PR   21-FEB-2002; 2002US-0357844P.
PR   15-MAR-2002; 2002US-0364134P.
PR   08-APR-2002; 2002US-0370144P.
PR   08-APR-2002; 2002US-0370206P.
PR   08-APR-2002; 2002US-0370247P.
PR   17-APR-2002; 2002US-0372794P.
PR   21-APR-2002; 2002US-0371679P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D,  Porter M,  Johnson K,  Higgs B,  Castle A,  Elashoff M;
XX
DR   WPI; 2003-148464/14.
XX
PT   Predicting at least one toxic effect of a compound, useful for toxicity
PT   modeling, comprises preparing a gene expression profile of a tissue or
PT   cell sample exposed to the compound, and comparing the gene expression
PT   profile to a database.
XX
PS   Example 4; Page; 446pp; English.
XX
CC   The invention relates to a novel method of predicting at least one toxic
CC   effect of a compound. The method comprises a gene expression profile of a
CC   tissue or cell sample exposed to the compound, and comparing the gene
CC   expression profile to a database comprising at least part of the data or
CC   information given in the specification. The methods are useful for
CC   predicting at least one toxic effect of a compound, predicting the
CC   progression of a toxic effect of a compound, predicting the renal
CC   toxicity of a compound, or identifying toxicity markers in tissues or
CC   cells exposed to known renal toxin. The genes are useful as toxicity
CC   markers in drug screening and toxicity assays, in monitoring disease or
CC   physiological states, or disease progression. This polynucleotide
CC   represents a rat DNA sequence relating to the toxic effect database
CC   described in the specification. NOTE: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the World Intellectual Property
CC   Organization
XX
SQ   Sequence 2481 BP; 599 A; 653 C; 555 G; 674 T; 0 U; 0 Other;

  Query Match            62.7%;  Score 18.8;  DB 10;  Length 2481;
  Best Local Similarity  76.7%;  Pred. No. 1.7e+02;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              |||  |||||||||| || | |||||| | |
Db       1077 CTGTGCCCCTCTCGTCTAGGGAGTTCAGGT 1048


RESULT 5
ADA02762
ID   ADA02762 standard; DNA; 23107 BP.
XX
AC   ADA02762;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human RUNX3 carcinoma associated gene, SEQ ID NO:1280.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1280; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;
```

```
  Query Match             62.7%;  Score 18.8;  DB 9;  Length 23107;
  Best Local Similarity   76.7%;  Pred. No. 2.1e+02;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||||   || || ||||| ||| |||||||
Db      14873 CTGGGAGCCCCTGGACTCCAGAATTCCGCT 14902


RESULT 6
ADB72500
ID   ADB72500 standard; DNA; 23107 BP.
XX
AC   ADB72500;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human Runx3 gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 328; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;

  Query Match             62.7%;  Score 18.8;  DB 10;  Length 23107;
  Best Local Similarity   76.7%;  Pred. No. 2.1e+02;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||||   || || ||||| ||| |||||||
Db      14873 CTGGGAGCCCCTGGACTCCAGAATTCCGCT 14902


RESULT 7
ADC85242
ID   ADC85242 standard; DNA; 23107 BP.
XX
AC   ADC85242;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human Runx3 genomic sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 28; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;

  Query Match             62.7%;  Score 18.8;  DB 10;  Length 23107;
  Best Local Similarity   76.7%;  Pred. No. 2.1e+02;
  Matches   23;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||||   || || ||||| ||| |||||||
Db      14873 CTGGGAGCCCCTGGACTCCAGAATTCCGCT 14902


RESULT 8
ADM74357
ID   ADM74357 standard; DNA; 23107 BP.
XX
AC   ADM74357;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Human carcinoma associated (CA) nucleic acid #13.
XX
KW   Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW   carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW   cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2004072154-A1.
XX
PD   15-APR-2004.
XX
PF   30-NOV-2001; 2001US-00997722.
XX
PR   22-DEC-2000; 2000US-00747377.
PR   02-MAR-2001; 2001US-00798586.
XX
PA   (MORR/) MORRIS D W.
```

PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2004-328562/30.
XX
PT New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
PS Claim 1; SEQ ID NO 28; 29pp; English.
XX
CC The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;

Query Match 62.7%; Score 18.8; DB 12; Length 23107;
Best Local Similarity 76.7%; Pred. No. 2.1e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||||   || || ||||| ||| |||||||
Db      14873 CTGGGAGCCCCTGGACTCCAGAATTCCGCT 14902
```

RESULT 9
AAZ27624/c
ID AAZ27624 standard; DNA; 3210 BP.
XX
AC AAZ27624;
XX
DT 20-DEC-1999 (first entry)
XX
DE Plasmid SPIF-1.
XX
KW Extracellular compartment modification; floral cell; self-compatibility;
KW pollen-pistil interaction; self-incompatibility; insect growth control;
KW plasmid SPF-1; GPIS363 gene; cysteine protease inhibitor gene; ss.
XX
OS Synthetic.
XX
PN WO9949063-A1.
XX
PD 30-SEP-1999.
XX
PF 19-MAR-1999; 99WO-CA000237.
XX
PR 20-MAR-1998; 98US-0078728P.
XX
PA (MIAC) CANADA MIN AGRIC & AGRI-FOOD CANADA.
XX
PI Robert LS, Gleddie S;
XX
DR WPI; 1999-591104/50.
XX
PT Protein expression in floral cells for peptide display, mediating plant
PT sterility, and modifying pollen-pistil interactions.
XX
PS Example 6; Page 103-104; 113pp; English.
XX
CC This sequence represents the plasmid SPIF-1, containing a fusion of the
CC Brassica napus GPIS363 gene to the Onchocerca protease inhibitor gene.
CC The invention relates to a method for modifying the extracellular
CC compartment of a floral cell of a plant, that comprises expressing a
CC construct comprising a gene of interest encoding a protein, fusion
CC protein or peptide, or a fragment of them, which is capable of modifying
CC the composition of the extracellular compartment of the floral cell and
CC altering either the function, use or development of the floral cell or
CC modifying the interaction of the floral cell with other cells, within an
CC anther or pistil cell. The method is used to modify pollen-pistil
CC interaction or function, which mediates, produces or prevents self-
CC compatibility, self-incompatibility, out- or in-crossing or combinations
CC of these. The method is also used for localizing proteins on the surface
CC of pollen for the purpose of peptide display. The protein localized on
CC the surface of the pollen may be an antibody or antigen or is a protein
CC that is effective in controlling insect growth, behaviour, feeding,
CC development or reproduction
XX
SQ Sequence 3210 BP; 1050 A; 626 C; 595 G; 939 T; 0 U; 0 Other;

Query Match 62.0%; Score 18.6; DB 2; Length 3210;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 CTGGACCCCTCTCGACTCGAGAGTT 25
              |||||||||| |||| | || ||||
Db       1232 CTGGACCCCTTTCGAATGGATAGTT 1208
```

RESULT 10
AAZ27623/c
ID AAZ27623 standard; DNA; 3459 BP.
XX
AC AAZ27623;
XX
DT 20-DEC-1999 (first entry)
XX
DE Plasmid SPF-1.
XX
KW Extracellular compartment modification; floral cell; self-compatibility;
KW pollen-pistil interaction; self-incompatibility; insect growth control;
KW plasmid SPF-1; GPIS363 gene; cysteine protease gene; ss.
XX
OS Synthetic.
XX
PN WO9949063-A1.
XX
PD 30-SEP-1999.
XX
PF 19-MAR-1999; 99WO-CA000237.
XX
PR 20-MAR-1998; 98US-0078728P.
XX
PA (MIAC) CANADA MIN AGRIC & AGRI-FOOD CANADA.

```
XX
PI   Robert LS,  Gleddie S;
XX
DR   WPI; 1999-591104/50.
XX
PT   Protein expression in floral cells for peptide display, mediating plant
PT   sterility, and modifying pollen-pistil interactions.
XX
PS   Example 5; Page 100-102; 113pp; English.
XX
CC   This sequence represents the plasmid SPF-1, containing a fusion of the
CC   Brassica napus GPIS363 gene to the Sitophilus cysteine protease gene. The
CC   invention relates to a method for modifying the extracellular compartment
CC   of a floral cell of a plant, that comprises expressing a construct
CC   comprising a gene of interest encoding a protein, fusion protein or
CC   peptide, or a fragment of them, which is capable of modifying the
CC   composition of the extracellular compartment of the floral cell and
CC   altering either the function, use or development of the floral cell or
CC   modifying the interaction of the floral cell with other cells, within an
CC   anther or pistil cell. The method is used to modify pollen-pistil
CC   interaction or function, which mediates, produces or prevents self-
CC   compatibility, self-incompatibility, out- or in-crossing or combinations
CC   of these. The method is also used for localizing proteins on the surface
CC   of pollen for the purpose of peptide display. The protein localized on
CC   the surface of the pollen may be an antibody or antigen or is a protein
CC   that is effective in controlling insect growth, behaviour, feeding,
CC   development or reproduction
XX
SQ   Sequence 3459 BP; 1082 A; 721 C; 678 G; 978 T; 0 U; 0 Other;

  Query Match             62.0%;  Score 18.6;  DB 2;  Length 3459;
  Best Local Similarity   84.0%;  Pred. No. 2.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTT 25
              ||||||||||  |||| | || ||||
Db       1232 CTGGACCCCTTTCGAATGGATAGTT 1208


RESULT 11
AAH51601
ID   AAH51601 standard; DNA; 319608 BP.
XX
AC   AAH51601;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human chromosome 13q31-q33 genomic nucleotide sequence.
XX
KW   sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
KW   biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058510-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-IB000435.
XX
PR   30-MAR-1999;   99US-0126903P.
PR   30-APR-1999;   99US-0131971P.
PR   30-APR-1999;   99US-0132065P.
PR   14-JUL-1999;   99US-0143928P.
PR   27-JUL-1999;   99US-0145915P.
PR   29-JUL-1999;   99US-0146452P.
PR   29-JUL-1999;   99US-0146453P.
PR   28-OCT-1999;   99US-0162288P.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D,  Blumenfeld M,  Chumakov I,  Bougueleret L,  Bihain B;
PI   Essioux L;
XX
DR   WPI; 2000-619082/59.
XX
PT   Polynucleotides comprising sequences from sbg1 and g35018 biallelic
PT   markers are used for genotyping and detecting schizophrenia or bipolar
PT   disorder and predisposition to these disorders.
XX
PS   Claim 1; Page 409-493; 737pp; English.
XX
CC   AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC   g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC   human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC   biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC   AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein
CC   products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018
CC   cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC   to isolate sbg1 cDNAs, while sbg1 exons from different primates are
CC   represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC   amplicons which comprise biallelic markers located on the chromosome
CC   13q31-q33 locus are represented in AAH51700 -AAH51817. Biallelic markers
CC   are represented in the sequences by degenerate/undefined base codes. PCR
CC   primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC   the invention. The biallelic marker containing nucleotide sequences are
CC   used to determine the identity of the nucleotide at a biallelic marker in
CC   a sample DNA sequence. The nucleotide sequences may be labelled and used
CC   for genotyping by determining the identity of a nucleotide at a Region D-
CC   related biallelic marker in a biological sample from single or multiple
CC   subjects. By determining the frequency of a biallelic marker in a
CC   population an association between a genotype and a trait, a haplotype and
CC   a trait and a phenotype and a trait can be detected. The sequences can be
CC   used to determine a predisoposition to or early onset of schizophrenia or
CC   bipolar disorder or a beneficial response to or side effects related to
CC   treatment against schizophrenia or bipolar disorder
XX
SQ   Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 0 U; 274 Other;

  Query Match             61.3%;  Score 18.4;  DB 3;  Length 319608;
  Best Local Similarity   78.6%;  Pred. No. 3.8e+02;
  Matches   22;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCG 28
              ||| |||| |||||| | |||| || ||
Db       8690 CTGAACCCATCTCGATTTGAGATTTACG 8717


RESULT 12
AAS09301
ID   AAS09301 standard; DNA; 319608 BP.
XX
AC   AAS09301;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Human schizophrenia associated gene g35030 and biallelic markers A1-A71.
XX
KW   Human; g35030; biallelic marker; A1-A71; chromosome 13q31-q33;
KW   schizophrenia; bipolar disorder; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   primer_bind      7938. .7958
FT                    /*tag=  a
FT                    /note= "Binds primer 99-27943.rp"
FT   primer_bind      8297. .8315
FT                    /*tag=  b
FT                    /note= "Binds primer 99-27943-150.mis"
FT   misc_binding     8304. .8328
FT                    /*tag=  c
FT                    /bound_moiety= "Probe_99-27943-150"
FT   misc_feature     8316
FT                    /*tag=  d
FT                    /note= "Biallelic marker A1"
```

```
FT   primer_bind     complement(8317. .8335)
FT                   /*tag= e
FT                   /note= "Binds primer 99-27943-150.mis complement"
FT   primer_bind     complement(8446. .8465)
FT                   /*tag= f
FT                   /note= "Binds primer 99-27943.pu complement"
FT   primer_bind     21365. .21385
FT                   /*tag= g
FT                   /note= "Binds primer 99-27935.rp"
FT   primer_bind     21653. .21671
FT                   /*tag= h
FT                   /note= "Binds primer 99-27935-193.mis"
FT   misc_binding    21660. .21684
FT                   /*tag= i
FT                   /bound_moiety= "Probe_99-27935-193"
FT   misc_feature    21672
FT                   /*tag= j
FT                   /note= "Biallelic marker A2"
FT   primer_bind     complement(21673. .21691)
FT                   /*tag= k
FT                   /note= "Binds primer 99-27935-193.mis complement"
FT   primer_bind     complement(21845. .21864)
FT                   /*tag= l
FT                   /note= "Binds primer 99-27935.pu complement"
FT   primer_bind     65463. .65471
FT                   /*tag= m
FT                   /note= "Binds primer 8-128.pu"
FT   primer_bind     65466. .65484
FT                   /*tag= n
FT                   /note= "Binds primer 8-128-33.mis"
FT   misc_binding    65473. .65497
FT                   /*tag= o
FT                   /bound_moiety= "Probe_8-128-33"
FT   misc_feature    65485
FT                   /*tag= p
FT                   /note= "Biallelic marker A3"
FT   primer_bind     complement(65486. .65504)
FT                   /*tag= q
FT                   /note= "Binds primer 8-128-33.mis complement"
FT   primer_bind     complement(65856. .65874)
FT                   /*tag= r
FT                   /note= "Binds primer 8-128.rp complement"
FT   primer_bind     95034. .95053
FT                   /*tag= s
FT                   /note= "Binds primer 99-31960.pu"
FT   primer_bind     95377. .95395
FT                   /*tag= t
FT                   /note= "Binds primer 99-31960-363.mis"
FT   misc_binding    95384. .95408
FT                   /*tag= u
FT                   /bound_moiety= "Probe_99-31960-363"
FT   misc_feature    95396
FT                   /*tag= v
FT                   /note= "Biallelic marker A4"
FT   primer_bind     complement(95397. .95415)
FT                   /*tag= w
FT                   /note= "Binds primer 99-31960-363.mis complement"
FT   primer_bind     complement(95543. .95563)
FT                   /*tag= x
FT                   /note= "Binds primer 99-31960.rp complement"
FT   primer_bind     107022. .107040
FT                   /*tag= y
FT                   /note= "Binds primer 99-24656.pu"
FT   primer_bind     107262. .107280
FT                   /*tag= z
FT                   /note= "Binds primer 99-24656-260.mis"
FT   misc_binding    107269. .107293
FT                   /*tag= aa
FT                   /bound_moiety= "Probe_99-24656-260"
FT   misc_feature    107281
FT                   /*tag= ab
FT                   /note= "Biallelic marker A5"
FT   primer_bind     complement(107282. .107300)
FT                   /*tag= ac
FT                   /note= "Binds primer 99-24656-260.mis complement"
FT   primer_bind     complement(107495. .107513)
FT                   /*tag= ad
FT                   /note= "Binds primer 99-24656.rp complement"
FT   primer_bind     160279. .160298
FT                   /*tag= ae
FT                   /note= "Binds primer 99-24639.rp"
FT   primer_bind     160621. .160639
FT                   /*tag= af
FT                   /note= "Binds primer 99-24639-163.mis"
FT   misc_binding    160628. .160652
FT                   /*tag= ag
FT                   /bound_moiety= "Probe_99-24639-163"
FT   misc_feature    160640
FT                   /*tag= ah
FT                   /note= "Biallelic marker A6"
FT   primer_bind     complement(160641. .160659)
FT                   /*tag= ai
FT                   /note= "Binds primer 99-24639-163.mis complement"
FT   primer_bind     160770. .160787
FT                   /*tag= aj
FT                   /note= "Binds primer 99-24634.pu"
FT   primer_bind     complement(160785. .160802)
FT                   /*tag= ak
FT                   /note= "Binds primer 99-24639.pu complement"
FT   primer_bind     160857. .160875
FT                   /*tag= al
FT                   /note= "Binds primer 99-24634-108.mis"
FT   misc_binding    160864. .160888
FT                   /*tag= am
FT                   /bound_moiety= "Probe_99-24634-108"
FT   misc_feature    160876
FT                   /*tag= an
FT                   /note= "Biallelic marker A7"
FT   primer_bind     complement(160877. .160895)
FT                   /*tag= ao
FT                   /note= "Binds primer 99-24634-108.mis complement"
FT   primer_bind     complement(161240. .161257)
FT                   /*tag= ap
FT                   /note= "Binds primer 99-24634.rp complement"
FT   primer_bind     168813. .168830
FT                   /*tag= aq
FT                   /note= "Binds primer 99-7652.pu"
FT   primer_bind     168955. .168973
FT                   /*tag= ar
FT                   /note= "Binds primer 99-7652-162.mis"
FT   misc_binding    168962. .168986
FT                   /*tag= as
FT                   /bound_moiety= "Probe_99-7652-162"
FT   misc_feature    168974
FT                   /*tag= at
FT                   /note= "Biallelic marker A8"
FT   primer_bind     complement(168975. .168993)
FT                   /*tag= au
FT                   /note= "Binds primer 99-7652-162.mis complement"
FT   primer_bind     complement(169331. .169351)
FT                   /*tag= av
FT                   /note= "Binds primer 99-7652.rp complement"
FT   primer_bind     170666. .170686
FT                   /*tag= aw
FT                   /note= "Binds primer 99-16100.pu"
FT   primer_bind     170791. .170809
FT                   /*tag= ax
FT                   /note= "Binds primer 99-16100-147.mis"
FT   misc_binding    170798. .170822
FT                   /*tag= ay
FT                   /bound_moiety= "Probe_99-16100-147"
FT   misc_feature    170810
FT                   /*tag= az
FT                   /note= "Biallelic marker A9"
FT   primer_bind     complement(170811. .170829)
FT                   /*tag= ba
```

```
FT                   /note= "Binds primer 99-16100-147.mis complement"
FT   primer_bind     complement(171153. .171173)
FT                   /*tag=  bb
FT                   /note= "Binds primer 99-16100.rp complement"
FT   primer_bind     173065. .173085
FT                   /*tag=  bc
FT                   /note= "Binds primer 99-5862.rp"
FT   primer_bind     173339. .173357
FT                   /*tag=  bd
FT                   /note= "Binds primer 99-5862-167.mis"
FT   misc_binding    173346. .173370
FT                   /*tag=  be
FT                   /bound_moiety= "Probe_99-5862-167"
FT   misc_feature    173358
FT                   /*tag=  bf
FT                   /note= "Biallelic marker A10"
FT   primer_bind     complement(173359. .173377)
FT                   /*tag=  bg
FT                   /note= "Binds primer 99-5862-167.mis complement"
FT   primer_bind     complement(173495. .173514)
FT                   /*tag=  bh
FT                   /note= "Binds primer 99-5862.pu complement"
FT   primer_bind     189753. .189771
FT                   /*tag=  bi
FT                   /note= "Binds primer 99-5919.pu"
FT   primer_bind     189938. .189956
FT                   /*tag=  bj
FT                   /note= "Binds primer 99-5919-215.mis"

  Query Match             61.3%;  Score 18.4;  DB 5;  Length 319608;
  Best Local Similarity   78.6%;  Pred. No. 3.8e+02;
  Matches   22;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTCCG 28
              ||| |||| |||||| | |||| || ||
Db       8690 CTGAACCCATCTCGATTTGAGATTTACG 8717


RESULT 13
AAL52198
ID   AAL52198 standard; cDNA; 340449 BP.
XX
AC   AAL52198;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Human secreted protein genomic DNA coding sequence.
XX
KW   Human; gene; ds; secreted protein; chromosome 5; tissue typing;
KW   secreted protein-related disease; transgenic animal; drug screening;
KW   pharmacogenomic analysis; single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(1107, T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5479, G)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5480, G)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6100, T)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6122, A)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6883, G)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7641, G)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8493, T)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9626, A)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9796, A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9868, A)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10209, C)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10311, G)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10831, T)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12230, C)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13378, C)
FT                   /*tag=  p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13423, T)
FT                   /*tag=  q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13425, G)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14116, G)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14508, C)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14645, A)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14932, A)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16272, G)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16945, T)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17065, G)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17442, G)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17475, C)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17844, A)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17929, G)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18540, C)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18592, T)
```

```
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18616, G)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20491, T)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22023, G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22248, G)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23118, C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23789, A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23800, A)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23930, A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24159, G)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26339, A)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26508, G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26570, A)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26582, G)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26693, A)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26884, T)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27320, T)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27339, A)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27542, G)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28586, A)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28591, T)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28599, A)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30857, A)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31503, T)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33671, A)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35045, C)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35944, G)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37157, G)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37603, T)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(39242, G)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(39404, A)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(40395, C)
FT                   /*tag= bj

  Query Match             60.7%;  Score 18.2;  DB 8;  Length 340449;
  Best Local Similarity   87.0%;  Pred. No. 4.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 TGGACCCCTCTCGACTCGAGAGT 24
              |||||||||| | |||| |||||
Db      25999 TGGACCCCTCCCCACTCCAGAGT 26021


RESULT 14
ADB74382
ID   ADB74382 standard; DNA; 42325 BP.
XX
AC   ADB74382;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Mycobacterium leprae DNA #16.
XX
KW   Non-naturally occurring peptide; anion pump protein; tuberculosis;
KW   hypersensitivity reaction; tuberculostatic; gene; ds.
XX
OS   Mycobacterium leprae.
XX
PN   US6583266-B1.
XX
PD   24-JUN-2003.
XX
PF   16-SEP-1994;   94US-00311731.
XX
PR   19-AUG-1993;   93US-00109181.
PR   22-OCT-1993;   93US-00142558.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Smith DR,  Mao J;
XX
DR   WPI; 2003-656441/62.
XX
PT   New Mycobacterium tuberculosis anion pump peptide useful for as
PT   tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
PS   Disclosure; SEQ ID NO 131; 26pp; English.
XX
CC   The invention relates to a non-naturally occurring peptide of
CC   Mycobacterium tuberculosis comprising an amino acid sequence
CC   corresponding to an anion pump protein. The invention also relates to a
CC   non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC   Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC   useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC   leprae or for screening for new tuberculosis drugs. Purified proteins
```

```
CC   derived from the sequences of the invention may elicit a specific immune
CC   response. The peptide may also be used to detect hypersensitivity
CC   reactions of individuals exposed to Mycobacterium tuberculosis or
CC   Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC   supports to detect antibodies typical of hypersensitivity reactions, from
CC   a patient's sera. This sequence represents Mycobacterium leprae DNA of
CC   the invention. Note: The sequence data for this patent did not form part
CC   of the printed specification but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 42325 BP; 9673 A; 13128 C; 11330 G; 8194 T; 0 U; 0 Other;

  Query Match             60.0%;  Score 18;  DB 10;  Length 42325;
  Best Local Similarity   80.8%;  Pred. No. 5e+02;
  Matches   21;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          5 ACCCCTCTCGACTCGAGAGTTCCGCT 30
              ||||| |||||||| | |||  ||||
Db        688 ACCCCACTCGACTCCACAGTGACGCT 713


RESULT 15
ADM27081_10/c
Continuation (11 of 17) of ADM27081 from base 1000001 (Hyperthermophile Methanopyrus kan
WP Sequence split into 17 fragments  LOCUS ADM27081 Accession Adm27081
WP     Fragment Name      Begin     End
WP     ADM27081_00            1   110000
WP     ADM27081_01       100001   210000
WP     ADM27081_02       200001   310000
WP     ADM27081_03       300001   410000
WP     ADM27081_04       400001   510000
WP     ADM27081_05       500001   610000
WP     ADM27081_06       600001   710000
WP     ADM27081_07       700001   810000
WP     ADM27081_08       800001   910000
WP     ADM27081_09       900001  1010000
WP     ADM27081_10      1000001  1110000
WP     ADM27081_11      1100001  1210000
WP     ADM27081_12      1200001  1310000
WP     ADM27081_13      1300001  1410000
WP     ADM27081_14      1400001  1510000
WP     ADM27081_15      1500001  1610000
WP     ADM27081_16      1600001  1694968

  Query Match             60.0%;  Score 18;  DB 11;  Length 110000;
  Best Local Similarity   80.8%;  Pred. No. 5.4e+02;
  Matches   21;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 CTGGACCCCTCTCGACTCGAGAGTTC 26
              ||   || |||| ||||||||||||||
Db      69684 CTTTGCCGCTCTGGACTCGAGAGTTC 69659


Search completed: April 26, 2005, 11:11:26
Job time : 433 secs
```